

## Textile-GAN

### — StyleGAN の潜在空間探索による印象制御に基づくテクスチャ柄生成 —

谷口 史果<sup>†</sup> 飛谷 謙介<sup>††</sup> 都賀 美有紀<sup>†</sup> 長田 典子<sup>†</sup>

<sup>†</sup>関西学院大学大学院理工学研究科 〒669-1330 兵庫県三田市学園上ヶ原 1

<sup>††</sup>長崎県立大学情報システム学部 〒851-2130 長崎県西彼杵郡長与町まなび野 1-1-1

E-mail: <sup>†</sup>{ixs50880, toga.m, nagata}@kwansei.ac.jp, <sup>††</sup>tobitani@sun.ac.jp

あらまし 衣服の柄 (テキスタイルデザイン) に関する感性的質感認知に着目し, StyleGAN とその潜在空間探索によるスタイルの制御を行うことにより, 所望の印象を持つテクスチャ柄の生成手法を提案する. 柄画像を対象とした主観評価実験で, 柄から喚起される印象に対する定量的な指標を作成する. これを学習データとして StyleGAN を構成し, 潜在空間探索によるスタイル制御を行うことで, モデルの生成画像の印象制御を実現する. 実験結果より, 生成画像の高精度な印象制御が確認され, 本手法の有効性が示された.

キーワード 感性工学, 生成 AI, StyleGAN, スタイル変換, 潜在空間, ファッション

## Textile-GAN

### — Textile-pattern generation based on impression control by latent space search in StyleGAN —

Fumika TANIGUCHI<sup>†</sup> Kensuke TOBITANI<sup>††</sup> Miyuki TOGA<sup>†</sup> and Noriko NAGATA<sup>†</sup>

<sup>†</sup>Graduate School of Science and Technology, Kwansei Gakuin University 1 Gakuenuegahara, Sanda-shi, Hyogo, 669-1330 Japan

<sup>††</sup>Faculty of Information Systems, University of Nagasaki 1-1-1, Manabino, Nagayo-cho, Nishisonogi-gun, Nagasaki, 851-2130 Japan

E-mail: <sup>†</sup>{ixs50880, toga.m, nagata}@kwansei.ac.jp, <sup>††</sup>tobitani@sun.ac.jp

**Abstract** We propose a method for generating textured patterns with desired impressions by controlling style using StyleGAN and its latent space search, focusing on sensory texture perception of clothing patterns (textile design). We created a quantitative index of the impression evoked by the pattern through subjective evaluation experiments on pattern images. Using this as training data, we constructed a StyleGAN and perform style control through a latent space search to realize impression control of the model's generated images. The experiment's results show that the impression control of the generated images is highly accurate, indicating this method's effectiveness.

**Keywords** Kansei Engineering, Generative AI, StyleGAN, Style Transformation, Latent Space, Fashion

#### 1. はじめに

感性的質感は, 良し悪しや好き嫌いに関わる価値付けを伴う質感と定義される. 本研究では感性の階層構造[1]で表される対象物の印象によって質感を評価する. プロダクトデザインの分野では個人のニーズや嗜好にあわせたビスポークデザイン (カスタマイゼーション) が注目されており, 感性的質感を理解し制御する技術の必要性が高まっている[2]. 特にファッション分野において, 衣服の柄 (パターンや配色などの模様) は, 衣服の選択に関わる主要な要素であり, 柄としてのテクスチャおよびその印象の対応関係の把握は重要である. これまでに柄に関するテクスチャの研究では,

画像特徴量の研究やシームレス化の研究, また最近では柄の好悪, 審美性 (aesthetics) 等の研究が盛んに行われている. しかし柄が持つ多様な印象や印象に起因する感情を体系的に説明し指標化する研究は, まだ十分に検討されていない.

一方, 最近の生成 AI (Generative AI) では, Stable Diffusion[3]モデルに代表されるようにテキストプロンプトから所望の画像を生成することが可能となっている. しかしながら, 柄の印象や感情といった心理的な指標に基づいた, 直感的かつ効率的なテクスチャの生成技術は確立されていない.

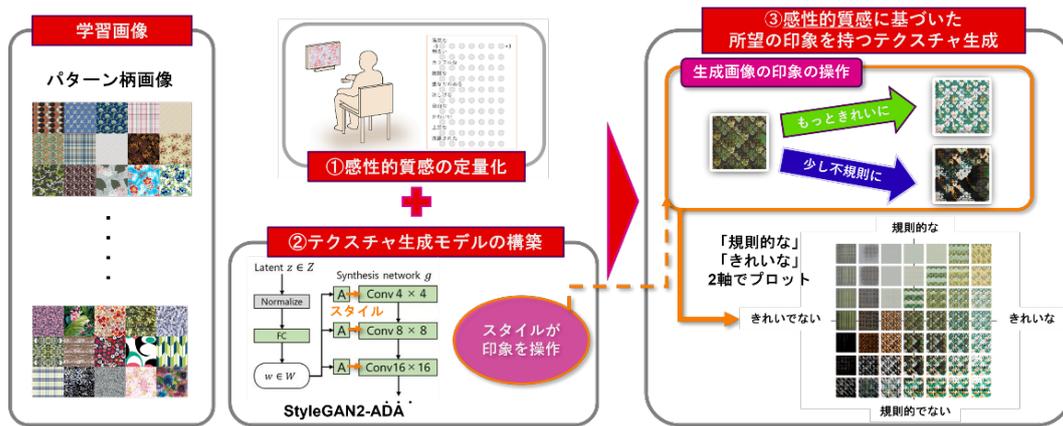


図 1 提案手法の概念図

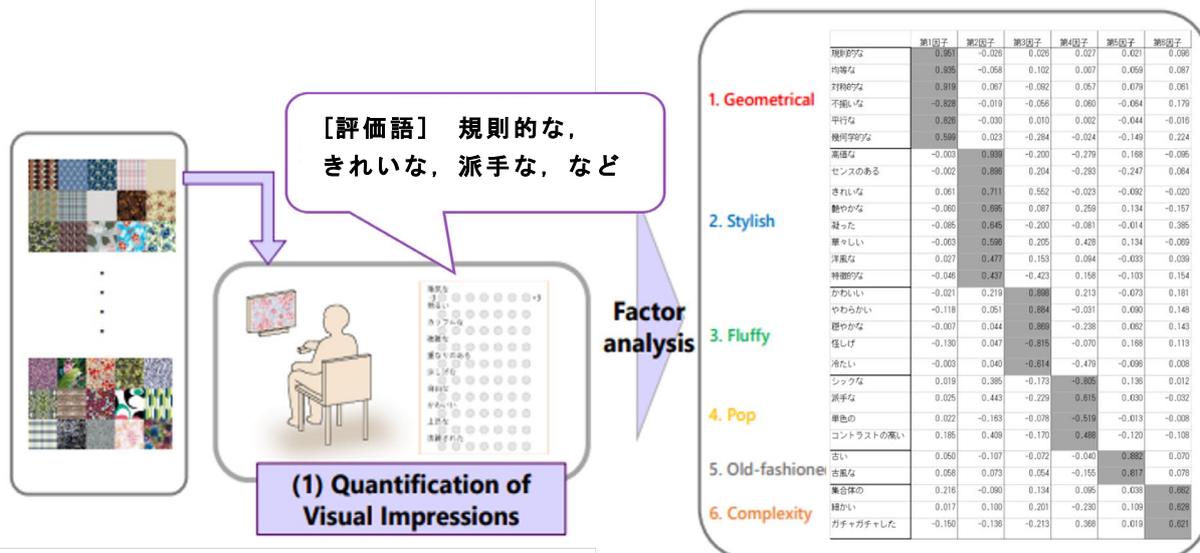


図 2 印象評価実験と因子分析

本研究では、衣服の柄から喚起される印象の定量的な評価を行い、これに基づき所望の印象を持つ柄（テキスト）を生成する手法として、StyleGAN の潜在空間探索を利用した方法を提案する。

## 2. 関連研究

画像生成モデルの代表的なものの1つに、敵対的生成ネットワーク（GAN）が挙げられる。とりわけ StyleGAN[4]は、画像のスタイル情報を含む潜在変数を利用して、より直感的に生成画像のスタイルを制御することが可能である。一方、衣服の柄に関する感性的質感に関する著者らの研究では、柄に関する人の評価構造・評価基準を明らかにし、画風変換アルゴリズム関連[5]のスタイル特徴を利用して、人の直感に上手く合致した柄検索機能を実現できることを示した[6]。しかし同じ枠組みによるスタイル特徴を用いた柄生成処理では、まだ十分な効果が得られていない。そこで本研究では、StyleGAN を用いた柄生成を試みる。

## 3. 提案手法

本研究では、StyleGAN とその潜在空間探索によるスタイルの制御を行うことにより、感性的質感認知に基づく柄生成を行う。柄画像を対象を絞った印象評価実験によって、柄画像に関する定量的な指標を得て、これを学習データとすることで高精度な印象制御能力を持つテキスト生成モデルの構築手法を提案する。StyleGAN の中でも少ない教師データ数での安定的な学習が可能である StyleGAN2-ada[7]を使用する。

具体的には、StyleGAN2-ada のマルチラベル学習を用いて所望の感性的質感を持つテキスト生成モデルを構築する。図 1 に提案手法の概念図を示す。はじめに、柄画像の感性的質感を定量化するための印象評価実験を行い、この実験で得られた結果を教師ラベルとして紐づけた柄画像をデータセットとして用いる。次に、これらのデータセットを用いて StyleGAN2-ada のマルチラベル学習を行う。以上の基本設計により、印象に対してもつれの解けた（disentangled）潜在空間を獲得し、入力画像と似た印象を持つテキストの生成や、生成画像の印象の操作が可能で、所望の感性的質

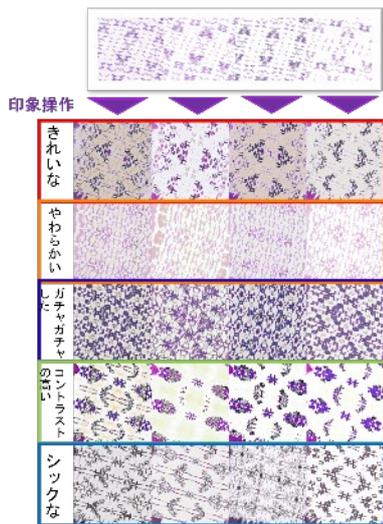


図3 評価語 28 語条件下の生成画像



図5 評価語 5~6 語条件下の生成画像

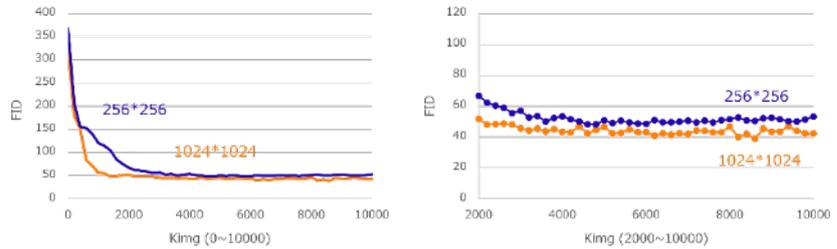


図6 学習画像の解像度ごとの FID 比較

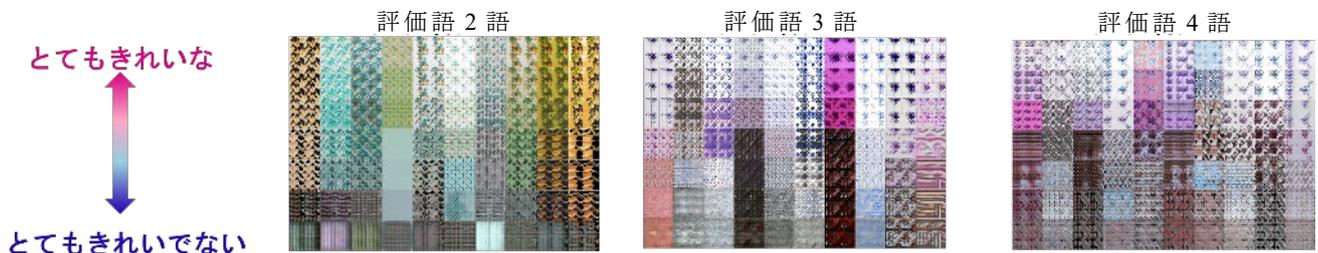


図4 評価語 2~4 語条件下の生成画像

感を持ったテクスチャの生成モデルを構築する。

### 3.1. 感性的質感の定量化

一般的な生成 AI モデルのような大量の web データで作成されたデータセットでは、明確な基準がないため、定量的な指標に基づいていないわけではない。本研究では柄から喚起される感性的質感に関する評価基準を、始めに構造化し、次に 7 件法に基づく主観評価実験を行い、その結果をデータセットとする。手順は以下の通りである。

衣服の柄画像 4036 枚を収集し、柄から喚起される感性的質感を表現する形容詞を自由記述方式で収集し、段階的に選択を進め、最終的に“きれいな”、“規則的な”、“派手な”、“集合体の”、のような計 28 語を選定した。その後、各画像に対し、選定した 28 語についての程度当てはまるかを -3 から 3 の 7 段階で評価する心理実験を行った[6]。この実験により得られた各柄画像に対する評価の平均値を、各画像の評価点とした。その後、図 2 のように因子分析によって「幾何学的」「スタイリッシュ」「複雑さ」といった 6 因子を抽出した。累積寄与率は 69.15%であった。

本研究では、この実験で用いた柄画像 4036 枚のデータセットと、各画像に付与された評価語ごとの評価点を学習に用い、感性的質感認知に基づく画像を生成

する。

### 3.2. テクスチャ生成モデルの構築

感性的質感の定量化で得られたデータセットを用いて StyleGAN2-ada でマルチラベル学習することでモデルを構築する。StyleGAN や StyleGAN2-ada のジェネレータは、学習された定数入力から開始し、潜在コードに基づいて各畳み込み層で画像の「スタイル」を調整する。スタイルは、ジェネレータの各畳み込み層で生成される画像の外観を制御する学習済みベクトルのことを指す。スタイルベクトルは学習プロセス中に学習され、正規分布からランダムにサンプリングされた高次元ベクトルである潜在コードに基づく。各畳み込み層でスタイルベクトルを調整することにより、ジェネレータは異なるスケールで画像特徴の強さを制御することができ、生成される画像を制御することができる。これにより高レベルの属性（柄のモチーフなどの大まかな特徴）と、生成された画像における確率的変動（柄の細かい特徴）を教師なしで分離が可能である。

データセットは柄画像と、前項の各柄画像に対応した評価語ごとの評価点である。学習画像の解像度は 256\*256pixel と 1024\*1024pixel を用いた。教師データである画像は 4036 枚と多くないが、それを解決するために StyleGAN2-ada を用いてテクスチャ生成モデル



図 7 各印象を高めた生成画像編集

(左から平均的な印象の生成画像，集合体の・派手な・きれいな・集合体の印象を高めた画像)

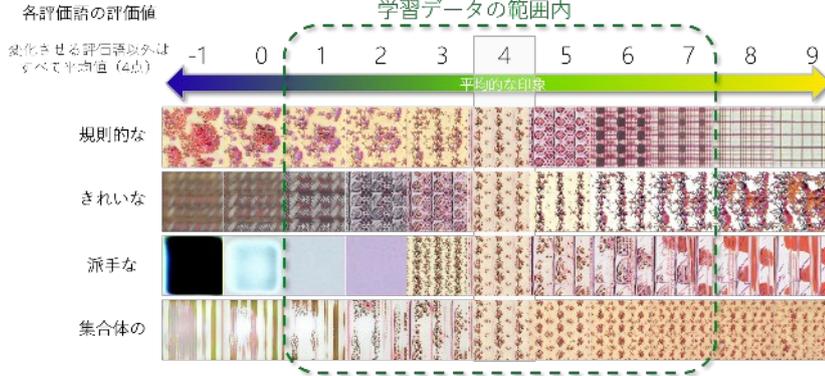


図 8 各印象に基づいた生成画像の印象制御結果

(内挿範囲は 1~7，外挿範囲は-1~0, 8~9)

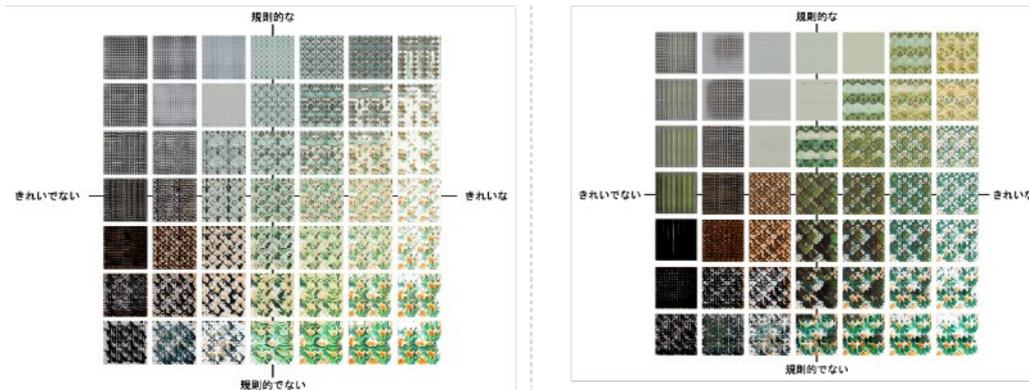


図 9 2つの生成画像の印象制御結果

を構築する。StyleGAN2-ada は適応的識別器増強メカニズム (adaptive discriminator augmentation, ADA) により安定的な学習が可能である。学習画像ごとに付与された各評価語 (印象) の評価点をラベルとして、柄の感性的質感をモデルにマルチラベル学習する。これにより、印象に対してもつれの解けた (disentangled) 潜在空間を獲得でき、評価語ごとの評価点を指定することで、定量的な指標に基づいたユーザ所望の印象を持つテクスチャの生成が可能になる。

学習条件は、以下のパラメータに着目し、最適な組み合わせを探索する。

- ・ R1 正則化 gamma(default:13 or 32 or 64)
- ・ batch サイズ(default:16 or 32)

## 4. 結果と考察

### 4.1. マルチラベル学習

はじめに StyleGAN2-ada のデフォルト設定で評価語

28 語すべてを同時にマルチラベル学習した。図 3 のように各印象は表現できていると考えられるが、指定するラベル、つまり評価語ごとの印象値が同一の場合、柄のモチーフがほぼ同様な画像が大量に生成された。この結果から 4036 枚の学習画像に対して 28 個のマルチラベル学習は、ラベル数が多すぎると考えられた。

そこで 6 因子の中から 1 つずつ評価語を選択し、実験した。例えば「Geometrical」因子からは「規則的な」の評価語を選択している。図 4 はラベル 2~4 個を学習させた生成画像 10 枚に対して、評価語 4 個の一つである「きれいな」の印象を変化させた結果である。学習した印象「きれいな」を制御し、表現できていると考えられる。図 5 はラベル 5~6 個を学習させた生成画像 10 枚であるが、同じ評価語の評価点を指定すると生成画像の多様性が失われた。図 4,5 はデフォルト設定でマルチラベル学習したが、各パラメータや損失関数を適切に設定することで 5 個以上のマルチラベル学習は可能になる可能性がある。本研究では

同時に「規則的な」「きれいな」「派手な」「集合体の」の4個の評価語をラベルとして同時に学習している。

#### 4.2. 解像度ごとの FID 推移

生成画像の評価を FID (Fréchet Inception Distance) 指標を用いて評価する。FID は生成画像の品質を評価するための一般的な指標であり [8], 事前に訓練された Inception-v3 ネットワークによって学習された特徴空間において、実画像の分布と生成画像の分布の間の距離を測定する。特徴空間における実画像と生成画像の分布を比較することで、ジェネレータがこれらの高レベルの特徴をどの程度正確に捉えることができるかを判断する定量的な尺度となる。FID スコアが低ければ低いほど、生成された画像は高レベルの特徴に関して実画像により類似していることを示す。これはジェネレータがより高品質な画像を生成していることを意味する。

解像度 256\*256pixel と 1024\*1024pixel の FID の推移は図 6 の通りである。この二つのパターンで比較すると、学習が安定してからの FID は大きな差があり、1024\*1024pixel の方が 256\*256pixel より FID が良かった。このことから学習柄画像の解像度が高いと生成画像の質も良くなることが示された。

#### 4.3. 評価指標の比較

FID の他に評価指標として Precision と Recall [9] も算出する。Precision は学習画像の画像特徴分布が生成画像のその分布を網羅している割合であり、正確性や精度に例えられる。それに対して Recall は、生成画像の分布が学習画像の分布を網羅している割合で、多様性に例えられる。この Precision と Recall はトレードオフであり、両方の値が高いモデルは十分に訓練されたモデルであると考え、最終的に複数のスナップショットのモデルから最良のものを選定する。

StyleGAN2-ada デフォルトの設定のまま学習したモデルと、FID, Precision, Recall で最良のスコアを記録したモデルを比較した。結果 (表 1) からわかる通り、R1 正則化 Gamma は 64 の条件、そして拡張方法は多い方が 3 つの指標のスコアが良かった。以降は、この Recall が最高のモデルを用いる。

#### 4.4. 生成画像の印象制御

ノイズからランダムに生成した画像 (解像度 1024\*1024pixel) の印象を編集した。4 つの評価語それぞれの印象を最大に高めた画像を生成した (図 7)。4 つの評価語に基づいて印象を段階的に変化させた結果は図 8 の通りであり、学習データ 1~7 の範囲内つまり内挿では変化が大きく評価語の印象通りに画像の印象を制御している様子が見て取れた。一方外挿に関しては、-1~9 の範囲内までなら変化が見られたが、それより外では大きな変化は見られなかった。また 4 つの評

価語の中で評価語“きれいな”と“規則的な”に基づいて印象を段階的に変化させ、その結果を 2 軸にプロットした (図 9)。

結果より、柄画像に対しての印象を表す 4 つの評価語の印象値を指定することで、定量的に所望の印象を持つ画像が生成されたことが確認された。

表 1 評価指標ごとのモデル比較

Model	FID	Recall	Precision
Gamma13 Batch32	39.8	0.24	0.65
Gamma32 Batch16	44.1	0.14	0.72
Gamma32 Batch32	39.6	0.17	0.72
Gamma64 Batch16	35.3	0.20	0.74
Gamma64 Batch32	40.8	0.26	0.74

表 2.1 生成画像の Lab 値の標準偏差比較

Model	FID	Recall	L値 標準偏差	a値 標準偏差	b値 標準偏差	L値 平均値	a値 平均値	b値 平均値
FID最高	35.3	0.20	26.6	13.0	14.6	53.3	0.09	1.14
Recall最高	40.8	0.26	27.1	13.1	18.5	54.7	0.50	0.73

### 5. 提案手法に対する妥当性検証

#### 5.1. Lab 値算出

すべての評価語の評価値を平均値の 4 点とした画像 1000 枚を生成し、生成画像の Lab 値を比較した。Lab 色空間では L 値は明度を表し値の範囲は 0~100 をとる。L 値は 1000 枚の平均値が 50 に近ければ生成画像の明度は偏っていないということになる。a 値は赤みの度合い、b 値は黄みの度合いを表し、a 値と b 値は 1000 枚の平均値が 0 に近ければ生成画像は偏った色相と彩度を持たない。

Precision 最高のモデルは Recall 最高のモデルと一致しているため、FID 最高のモデルと Precision・Recall 最高のモデルの 2 つを比較した。表 2 の通り、2 つのモデルは Lab 値の標準偏差に大きな違いがあり、平均値はさほど大きな違いが見られなかった。Recall が最高のモデルの方が Lab 値の標準偏差が高く、幅広い色合いのテクスチャを生成するということがわかった。図 10 の通り、生成画像の多様性はそれぞれのモデルにあるように考えられる。



(a) Recall と Precision 最高 (b) FID 最高

図 10 生成画像の抜粋と比較

## 5.2. 柄画像の評価値検証

寸田らが構築した柄テクスチャ印象推定モデル[6]を用いて、生成画像の評価値を算出することで、評価値通りに生成しているかを検証する。そのためにまず、正解データが付与されたデータセットの評価値を算出し、モデルの精度を確かめる。このプログラムは感性的質感とテクスチャのスタイル特徴を Random Forest 回帰によって定式化している。

現在、このモデルを用いてあまり高い相関は見られないため、今後は異なる方法を用いて生成柄画像の評価値検証を行う予定である。

## 5.3. 生成画像のクラスタリング

5.2 では生成画像の多様性を見るために、Lab 値を算出した。しかし、生成画像の色相の情報は確認できるが柄のモチーフが同一で色や細部を変えた画像が散見された。そのため、次は生成画像の柄の多様性を確認するため、同系統の柄がどの程度の頻度で生成され、同系統の柄の中でもどの程度の差異があるかをクラスタリングで検証する。すべての評価語の評価値を平均値の4点とした画像を1000枚生成し、k-means法でクラスタリングする。

ResNet50モデルの最上位分類層を除外し、画像の特徴マップを出力し、その特徴マップから平均プーリングを用いて特徴ベクトルにした。各画像の特徴ベクトルをk-means法でクラスタリングした。この時のクラスタ数はエルボー法を用いた。図12の通りクラスタ内誤差平方和は緩やかに下がったことから多様な画像を生成していることが確認された。クラスタ数は50とした。クラスタリング結果の抜粋を図13に表す。概ね柄画像を色ではなくモチーフによって分類していることが視覚的に分かった。

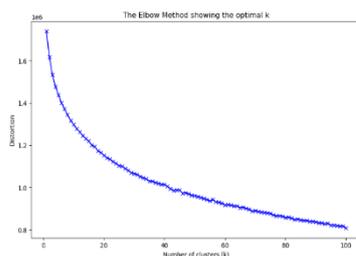


図12 エルボー法によるクラスタ数決定

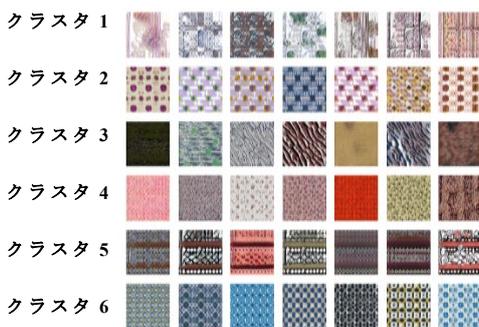


図13 クラスタリング結果の抜粋

## 6. まとめ

本研究では、衣服の柄に関する感性的質感認知に着目し、所望の印象を持つテクスチャ柄の生成手法を提案した。柄画像を対象とした主観評価実験で、柄から喚起される印象に対する定量的な指標を作成する。これを学習データとして StyleGAN を構成し、潜在空間探索によるスタイル制御を行うことで印象制御を実現した。実験結果より、生成画像の高精度な印象制御が確認され、本手法の有効性が示された。

今後は、テクスチャ生成モデルが同時に学習できる最大ラベル数を適切なパラメータ設定によって増やし実験するとともに、モデルが評価語とその評価値を正解データ通りに学習しているかを検証する。そのために、ラベルである各評価語の評価点を変化させた生成画像の潜在空間内の分布解析を検討している。

## 文 献

- [1] 片平建史, 武藤和仁, 橋本 翔, 飛谷謙介, 長田典子, “SD 法 を用いた感性的の測定における評価の階層性,” 日本感性工学会論文誌, vol. 17, no. 4, pp. 453-463, 2018.
- [2] 飛谷謙介, 白岩史, 片平建史, 長田典子, 荷方邦夫, 荒川薫, “化粧品パッケージデザインにおける高級感印象のモデル化,” 精密工学会誌, vol. 87, no. 1, pp. 134-139, 2021.
- [3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models”, Proceedings of the IEEE/CVF Conference on CVPR, pp. 10684-10695, Jun.2022.
- [4] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” Proceedings of ICLR 2018, May. 2018.
- [5] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image Style Transfer Using Convolutional Neural Networks,” International Conference on Computer Vision and Pattern Recognition, pp. 2414–2423, Jun.2016.
- [6] N. Sunda, K. Tobitani, I. Tani, Y. Tani, N. Nagata, and N. Morita, “Impression Estimation Model for Clothing Patterns Using Neural Style Features,” International Conference on Human-Computer Interaction, pp. 689– 697, Jul.2020.
- [7] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, “Training generative adversarial networks with limited data,” Advances in Neural Information Processing Systems, no. 1015, pp. 12104–12114, Dec.2020.
- [8] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, vol.30, pp. 6629-6640, Dec. 2017.
- [9] T. Kynkäänniemi, T. Karras, S. Laine, and J. Lehtinen, “Improved precision and recall metric for assessing generative models,” NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems, vol.33, pp. 3927-3936, Dec. 2019.